Dropout

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

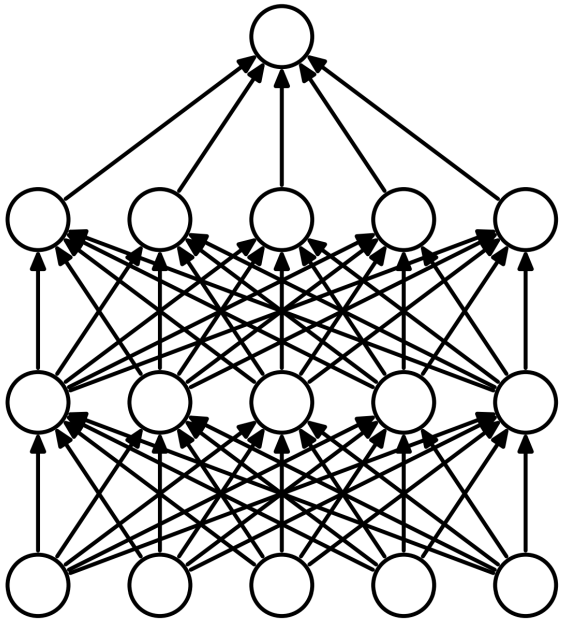
Review: Data Augmentation

- Simple idea:
  - 입력 데이터에 noise를 추가하자
  - 쓸모 없는 특징까지 배우지 않도록 하기 위함

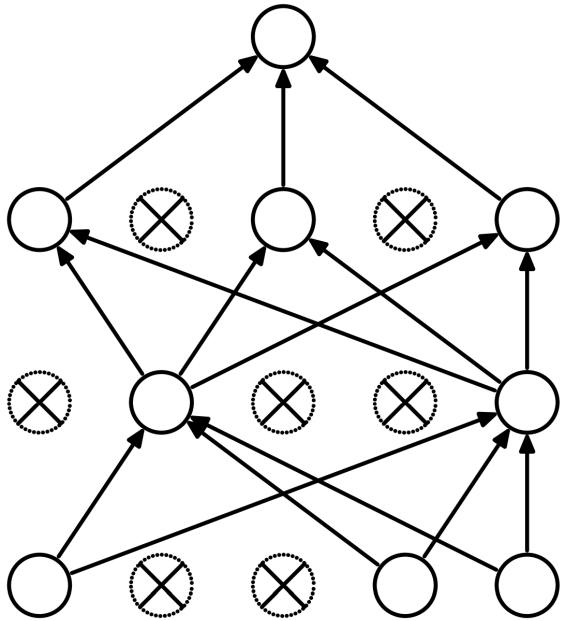
- 신경망 중간에 noise를 추가할 수는 없을까?

Dropout [Srivastava et al., 2014]

- 임의(e.g. 동전 던지기)로 노드를 동작하지 않도록 함



(a) Standard Neural Net

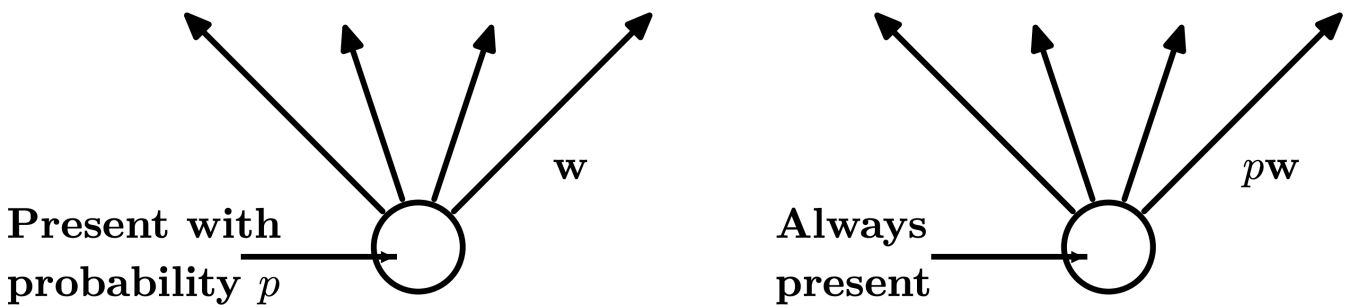


(b) After applying dropout.

학습과 추론의 방법이 다름

- 학습
  - 확률 p에 따라 노드를 turn-off

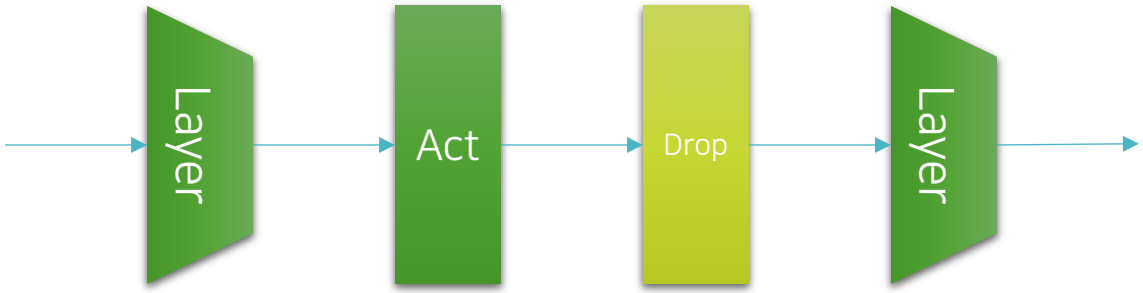
- 추론
  - 모든 노드를 turn-on
  - 하지만 학습 때보다 평균적으로 $\frac{1}{p}$ 배 더 큰 입력을 받게 될 것
  - 따라서 p를 W에 곱하여 이를 상쇄



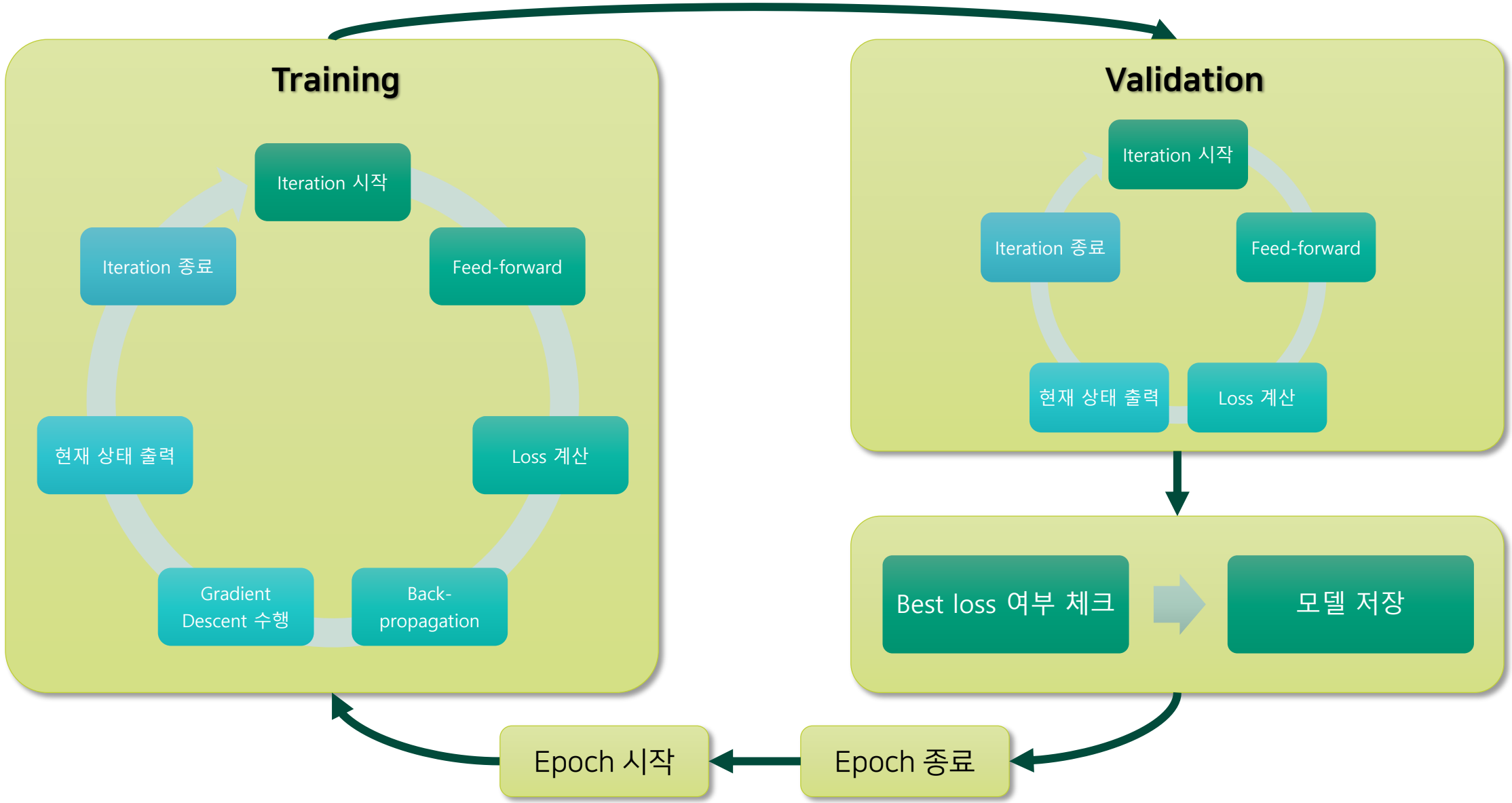
(a) At training time (b) At test time

Where?

- 레이어 → 활성함수 → **Dropout** → 레이어
  - 하이퍼 파라미터 p와 함께



Review: Typical Model Training Procedure



Model Mode 전환

- Loop for n_epochs:
  - Loop for n_mini_batches in train-set:
    - Feed-forward.
    - Calculate loss.
    - Back-propagation.
    - 1 step gradient descent.

  Training Iterations

  - Loop for n_mini_batches in validation-set:
    - Feed-forward.
    - Calculate loss.

  Validation Iterations

  - Save model if it is best.

Model Mode 전환

- Loop for n_epochs:
  - **model.train()**
  - Loop for n_mini_batches in train-set:
    - Feed-forward.
    - Calculate loss.
    - Back-propagation.
    - 1 step gradient descent.
  - **model.eval()**
  - Loop for n_mini_batches in validation-set:
    - Feed-forward.
    - Calculate loss.
  - Save model if it is best.

Training Iterations

Validation Iterations

참고: model은 training mode가 default setting.

Wrap-up

- Dropout을 통해 noise를 주어 학습을 방해

- 모델의 모드 전환을 잊지 말자!

- Pros:
  - Generalization error 감소
- Cons:
  - 학습 속도 저하
  - Hyper-parameter 추가: drop 확률 p